MULTI-DEVICE SPEECH RECOGNITION

BACKGROUND

[0001] Many modern devices support speech recognition. A significant limiting factor in utilizing speech recognition is the quality of the audio sample. Among the factors that contribute to low or diminished quality audio samples are background noise and movement of the speaker in relation to the audio capturing device.

[0002] One approach to improving the quality of an audio sample is to utilize an array of microphones. Often, however, a microphone array will need to be calibrated to a specific setting before it can be effectively utilized. Such a microphone array is not well suited for a user that frequently moves from one setting to another.

SUMMARY

[0003] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0004] In some embodiments, one or more secondary devices in physical proximity to a user of a principal device may be identified. Each of the secondary devices may be configured to capture audio. Multiple audio samples captured by the identified devices may be received. An audio sample comprising a voice of the user of the principal device may be selected from among the audio samples captured by the secondary devices based on suitability of the audio sample for speech recognition.

[0005] In some embodiments, the audio samples may be converted, via speech recognition, to corresponding text strings. Recognition confidence values corresponding to a level of confidence that a corresponding text string accurately reflects content of the audio sample from which it was converted may be determined. A recognition confidence value indicating a level of confidence as great or greater than the determined recognition confidence values may be identified, and an audio sample corresponding to the identified recognition confidence value may be selected. Additionally or alternatively, the audio samples may be analyzed to identify an audio sample that is equally well suited or more well suited for speech recognition and the identified audio sample may be selected.

[0006] In some embodiments, the audio samples captured by the secondary devices may include an audio sample comprising a voice other than the voice of the user of the principal device. The audio sample comprising the voice other than the voice of the user of the principal device may be identified by comparing each of the audio samples captured by the secondary devices to a reference audio sample of the voice of the user of the principal device. Once identified, the audio sample comprising the voice other than the voice of the user of the principal device may be discarded. Additionally or alternatively, the audio samples captured by the secondary devices may include an audio sample comprising both the voice of the user of the principal device and a voice other than the voice of the user of the principal device. The audio sample comprising both the voice of the user of the principal device and the voice other than the voice of the user of the principal device may be separated into two portions by comparing the audio sample comprising both the voice of the user of the principal device and the voice other than the voice of the user of the principal device to a reference audio sample of the voice of the user of the principal device. The first portion may comprise the voice of the user of the principal device and the second portion may comprise the voice of the user other than the user of the principal device. The second portion may be discarded.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The foregoing summary, as well as the following detailed description of illustrative embodiments, may be better understood when read in conjunction with the accompanying drawings, which are included by way of example, and not by way of limitation.

[0008] FIG. **1** illustrates an exemplary environment for multi-device speech recognition in accordance with one or more embodiments.

[0009] FIG. **2** illustrates an exemplary sequence for multi-device speech recognition in accordance with one or more embodiments.

[0010] FIG. **3** illustrates an exemplary method for selecting an audio sample based on a confidence level that a text string converted from the audio sample accurately reflects the content of the audio sample.

[0011] FIG. **4** illustrates an exemplary method for selecting an audio sample based on analyzing the suitability of the audio sample for speech recognition.

[0012] FIG. **5** illustrates an exemplary method for selecting an audio sample by dividing corresponding audio samples into multiple frames, selecting preferred frames based on their suitability for speech recognition, and combining the preferred frames to form a hybrid sample.

[0013] FIG. **6** illustrates an exemplary apparatus for multi-device speech recognition in accordance with one or more embodiments.

[0014] FIG. **7** illustrates an exemplary method for multi-device speech recognition.

DETAILED DESCRIPTION

[0015] FIG. **1** illustrates an exemplary environment for multi-device speech recognition in accordance with one or more embodiments. Referring to FIG. **1**, environment **100** may include user **102** and principal device **104**. Principal device **104** may be any device capable of utilizing a text string produced via speech recognition. For example, principal device **104** may be a smartphone, tablet computer, laptop computer, desktop computer, or other similar device capable of utilizing a text string produced via speech recognition. Environment **100** may also include secondary devices **106-118**. Secondary devices **106 -118** may include one or more devices capable of capturing audio associated with a user of principal device **104** (e.g., user **102**). For example, secondary devices **106-118** may include smartphones, tablet computers, laptop computers, desktop computers, speakerphones, headsets, microphones integrated into a room or vehicle, or any other device capable of capturing audio associated with a user of principal device **104**. As used herein, "principal device" refers to a device that utilizes output produced from an audio sample (e.g., a text string produced via speech recognition), and "secondary device" refers to any device, other than the principal device, that is capable of capturing audio associated with a user of the principal device. A principal device or a